Submitted to the Annals of Statistics

SUPPLEMENT TO "GEE ANALYSIS OF CLUSTERED BINARY DATA WITH DIVERGING NUMBER OF COVARIATES"'

By Lan Wang,

*University of Minnesota**

Throughout the proof, we use C to denote a genetic positive constant, which may vary from line to line.

APPENDIX A

Proof of (3.4). We will first show that

$$||\widehat{\mathbf{R}} - \mathbf{R}^*|| = O_p(\sqrt{p_n/n}), \tag{A.1}$$

where

$$\mathbf{R}^* = \frac{1}{n}\sum_{i=1}^n \mathbf{A}_i^{-1/2}(\boldsymbol{\beta}_{n0})(\mathbf{Y}_i - \boldsymbol{\pi}_i(\boldsymbol{\beta}_{n0}))(\mathbf{Y}_i - \boldsymbol{\pi}_i(\boldsymbol{\beta}_{n0}))^T\mathbf{A}_i^{-1/2}(\boldsymbol{\beta}_{n0}).$$

The central limit theorem implies that $||\mathbf{R}^* - \mathbf{R}_0|| = O_p(1/\sqrt{n})$. Combined with (A.1), this would imply that $||\widehat{\mathbf{R}} - \mathbf{R}_0|| = O_p(\sqrt{p_n/n})$. Note that $||\widehat{\mathbf{R}} - \mathbf{R}^*||^2 = \sum_{k=1}^m \sum_{j=1}^m [\widehat{\mathbf{R}}_{kj} - \mathbf{R}^*_{kj}]^2$, where

$$\widehat{\mathbf{R}}_{kj} - \mathbf{R}^*_{kj} = \frac{1}{n}\sum_{i=1}^n \left[\frac{(Y_{ik} - \widetilde{\pi}_{ik})(Y_{ij} - \widetilde{\pi}_{ij})}{\sqrt{\widetilde{\mathbf{A}}_{ik}}\sqrt{\widetilde{\mathbf{A}}_{ij}}} - \frac{(Y_{ik} - \pi^0_{ik})(Y_{ij} - \pi^0_{ij})}{\sqrt{\mathbf{A}^0_{ik}}\sqrt{\mathbf{A}^0_{ij}}}\right],$$

with $\pi^0_{ik} = \pi_{ik}(\boldsymbol{\beta}_{n0})$, $\widetilde{\pi}_{ik} = \pi_{ik}(\widetilde{\boldsymbol{\beta}}_n)$, $\mathbf{A}^0_{ik} = \pi^0_{ik}(1 - \pi^0_{ik})$ and $\widetilde{\mathbf{A}}_{ik} = \widetilde{\pi}_{ik}(1 - \widetilde{\pi}_{ik})$. We have

$$\begin{aligned}
|\widehat{\mathbf{R}}_{kj} - \mathbf{R}^*_{kj}| &\le \left|\frac{1}{n}\sum_{i=1}^n \frac{(Y_{ik} - \widetilde{\pi}_{ik})(Y_{ij} - \widetilde{\pi}_{ij}) - (Y_{ik} - \pi^0_{ik})(Y_{ij} - \pi^0_{ij})}{\sqrt{\mathbf{A}^0_{ik}}\sqrt{\mathbf{A}^0_{ij}}}\right| \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^n \frac{(Y_{ik} - \widetilde{\pi}_{ik})(Y_{ij} - \widetilde{\pi}_{ij})}{\sqrt{\mathbf{A}^0_{ik}}\sqrt{\mathbf{A}^0_{ij}}}\delta_{ijk}\right| \\
&\triangleq I_{kj,1} + I_{kj,2},
\end{aligned}$$

where $\delta_{ijk} = [\mathbf{A}_{ik}^0 \mathbf{A}_{ij}^0]^{1/2} [\widetilde{\mathbf{A}}_{ik} \widetilde{\mathbf{A}}_{ij}]^{-1/2} - 1$. Thus

$$||\widehat{\mathbf{R}} - \mathbf{R}^*||^2 \le 2 \sum_{k=1}^m \sum_{j=1}^m I_{kj,1}^2 + 2 \sum_{k=1}^m \sum_{j=1}^m I_{kj,2}^2 \quad \triangleq \quad I_{n1} + I_{n2}.$$

By triangle inequality, it is easy to see that

$$\begin{aligned} I_{kj,1} \quad \le & \quad \frac{1}{n} \sum_{i=1}^n \frac{|(\pi_{ik}^0 - \widetilde{\pi}_{ik})(\pi_{ij}^0 - \widetilde{\pi}_{ij})|}{\sqrt{\mathbf{A}_{ik}^0} \sqrt{\mathbf{A}_{ij}^0}} + \frac{1}{n} \sum_{i=1}^n \frac{|(\pi_{ik}^0 - \widetilde{\pi}_{ik})(Y_{ij} - \pi_{ij}^0)|}{\sqrt{\mathbf{A}_{ik}^0} \sqrt{\mathbf{A}_{ij}^0}} \\ & + \frac{1}{n} \sum_{i=1}^n \frac{|(\pi_{ij}^0 - \widetilde{\pi}_{ij})(Y_{ik} - \pi_{ik}^0)|}{\sqrt{\mathbf{A}_{ik}^0} \sqrt{\mathbf{A}_{ij}^0}} \triangleq I_{kj,11} + I_{kj,12} + I_{kj,13}. \end{aligned}$$

Thus

$$I_{n1} \le 6 \sum_{k=1}^m \sum_{j=1}^m I_{kj,11}^2 + 6 \sum_{k=1}^m \sum_{j=1}^m I_{kj,12}^2 + 6 \sum_{k=1}^m \sum_{j=1}^m I_{kj,13}^2 \triangleq I_{n11} + I_{n12} + I_{n13}.$$

By Cauchy-Schwarz inequality,

$$I_{kj,11}^2 \le \left[\frac{1}{n} \sum_{i=1}^n \frac{(\pi_{ik}^0 - \widetilde{\pi}_{ik})^2}{\pi_{ik}^0 (1 - \pi_{ik}^0)} \right] \left[\frac{1}{n} \sum_{i=1}^n \frac{(\pi_{ij}^0 - \widetilde{\pi}_{ij})^2}{\pi_{ij}^0 (1 - \pi_{ij}^0)} \right].$$

Therefore,

$$I_{n11} \le 6 \left[\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^m \frac{(\pi_{ik}^0 - \widetilde{\pi}_{ik})^2}{\pi_{ik}^0 (1 - \pi_{ik}^0)} \right] \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \frac{(\pi_{ij}^0 - \widetilde{\pi}_{ij})^2}{\pi_{ij}^0 (1 - \pi_{ij}^0)} \right].$$

Since $\frac{\partial}{\partial \boldsymbol{\beta}_n} \pi_{ik} = \pi_{ik}(\boldsymbol{\beta}_n)(1 - \pi_{ik}(\boldsymbol{\beta}_n)) \mathbf{X}_{ik}$, for some $\overline{\boldsymbol{\beta}}_n$ between $\widetilde{\boldsymbol{\beta}}_n$ and $\boldsymbol{\beta}_{n0}$,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^m \frac{(\pi_{ik}^0 - \widetilde{\pi}_{ik})^2}{\pi_{ik}^0 (1 - \pi_{ik}^0)} \\ = & \quad \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^m \frac{(\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0})^T \mathbf{X}_{ik} \mathbf{X}_{ik}^T (\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}) \pi_{ik}^2(\overline{\boldsymbol{\beta}}_n)(1 - \pi_{ik}^2(\overline{\boldsymbol{\beta}}_n))^2}{\pi_{ik}^0 (1 - \pi_{ik}^0)} \\ \le & \quad n^{-1} C (\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0})^T \left(\sum_{i=1}^n \mathbf{X}_i^T \mathbf{X}_i \right) (\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}) \\ \le & \quad n^{-1} C ||\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}||^2 \lambda_{\max} \left(\sum_{i=1}^n \mathbf{X}_i^T \mathbf{X}_i \right) = O(||\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}||^2), \end{aligned}$$

where the last inequality uses (A3). Thus $I_{n11} = O_p(p_n^2/n^2)$. Similarly, we can show that $I_{n12} = O_p(p_n/n)$ and $I_{n13} = O_p(p_n/n)$. Next, note that

$$\begin{aligned}
I_{n2} &\leq 2\sum_{k=1}^{m}\sum_{j=1}^{m}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{(Y_{ik}-\widetilde{\pi}_{ik})^2(Y_{ij}-\widetilde{\pi}_{ij})^2}{\pi_{ik}^0(1-\pi_{ik}^0)\pi_{ij}^0(1-\pi_{ij}^0)}\right]\left[\frac{1}{n}\sum_{i=1}^{n}\delta_{ijk}^2\right] \\
&\leq \frac{C}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}\sum_{j=1}^{m}\delta_{ijk}^2.
\end{aligned}$$

To evaluate δ_{ijk}, Let $g(\boldsymbol{\beta}_n) = [\mathbf{A}_{ik}^0\mathbf{A}_{ij}^0]^{1/2}[\mathbf{A}_{ik}(\boldsymbol{\beta}_n)\mathbf{A}_{ij}(\boldsymbol{\beta}_n)]^{-1/2}$, then $\delta_{ijk} = g(\widetilde{\boldsymbol{\beta}}_n) - g(\boldsymbol{\beta}_{n0})$. Note that $\frac{\partial}{\partial\boldsymbol{\beta}_n}g(\boldsymbol{\beta}_n) = g_{ijk,1}(\boldsymbol{\beta}_n)\mathbf{X}_{ik} + g_{ijk,2}(\boldsymbol{\beta}_n)\mathbf{X}_{ij}$, where

$$\begin{aligned}
g_{ijk,1}(\boldsymbol{\beta}_n) &= -\frac{1}{2}[\mathbf{A}_{ik}^0\mathbf{A}_{ij}^0]^{1/2}[\mathbf{A}_{ik}(\boldsymbol{\beta}_n)\mathbf{A}_{ij}(\boldsymbol{\beta}_n)]^{-1/2}(1-2\pi_{ik}(\boldsymbol{\beta}_n)), \\
g_{ijk,2}(\boldsymbol{\beta}_n) &= -\frac{1}{2}[\mathbf{A}_{ik}^0\mathbf{A}_{ij}^0]^{1/2}[\mathbf{A}_{ik}(\boldsymbol{\beta}_n)\mathbf{A}_{ij}(\boldsymbol{\beta}_n)]^{-1/2}(1-2\pi_{ij}(\boldsymbol{\beta}_n)).
\end{aligned}$$

Both $g_{ijk,1}(\boldsymbol{\beta}_n)$ and $g_{ijk,2}(\boldsymbol{\beta}_n)$ are bounded in probability. Thus for some $\boldsymbol{\beta}^*$ lies between $\widetilde{\boldsymbol{\beta}}_n$ and $\boldsymbol{\beta}_{n0}$, we have

$$\begin{aligned}
& I_{n2} \\
\leq\; & \frac{C}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}\sum_{j=1}^{m}(\widetilde{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_{n0})^T[g_{ijk,1}(\boldsymbol{\beta}^*)\mathbf{X}_{ik}+g_{ijk,2}(\boldsymbol{\beta}^*)\mathbf{X}_{ij}] \\
& [g_{ijk,1}(\boldsymbol{\beta}^*)\mathbf{X}_{ik}^T+g_{ijk,2}(\boldsymbol{\beta}^*)\mathbf{X}_{ij}^T](\widetilde{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_{n0}) \\
\leq\; & \frac{2C}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}\sum_{j=1}^{m}(\widetilde{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_{n0})^T[g_{ijk,1}^2(\boldsymbol{\beta}^*)\mathbf{X}_{ik}\mathbf{X}_{ik}^T+g_{ijk,2}^2(\boldsymbol{\beta}^*)\mathbf{X}_{ij}\mathbf{X}_{ij}^T](\widetilde{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_{n0}) \\
\leq\; & C||\widetilde{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_{n0}||^2[\lambda_{\max}(n^{-1}\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{X}_i)+\lambda_{\max}(n^{-1}\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{X}_i)] = O_p(p_n/n),
\end{aligned}$$

by assumption A3. Summarizing the above, (A.1) holds. Finally, (3.4) is proved by observing that $||\widehat{\mathbf{R}}^{-1}-\mathbf{R}_0^{-1}|| = ||\widehat{\mathbf{R}}^{-1}(\widehat{\mathbf{R}}-\mathbf{R}_0)\mathbf{R}_0^{-1}||$. $\square$

PROOF OF LEMMA 3.5.

$$
\begin{aligned}
& \mathbf{b}_n^T[\overline{\mathbf{H}}_n(\boldsymbol{\beta}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_{n0})]\mathbf{b}_n \\
= & \Big|\sum_{i=1}^n \mathbf{b}_n^T\mathbf{X}_i^T[\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n) - \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})]\mathbf{X}_i\mathbf{b}_n\Big| \\
\leq & \Big|\sum_{i=1}^n \mathbf{b}_n^T\mathbf{X}_i^T[\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n) - \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})]\mathbf{X}_i\mathbf{b}_n\Big| \\
& +\Big|\sum_{i=1}^n \mathbf{b}_n^T\mathbf{X}_i^T[\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0}) - \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})]\mathbf{X}_i\mathbf{b}_n\Big| \\
\triangleq & I_{n1} + I_{n2}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
I_{n1} & \leq \sum_{i=1}^n \Big|\mathbf{b}_n^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}[\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n) - \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})]\mathbf{X}_i\mathbf{b}_n\Big| \\
& \leq \sum_{i=1}^n ||\mathbf{b}_n^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1/2}|| \cdot ||\overline{\mathbf{R}}^{-1/2}[\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n) - \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})]\mathbf{X}_i\mathbf{b}_n||,
\end{aligned}
$$

by Cauchy-Schwarz inequality. We have

$$
\begin{aligned}
||\mathbf{b}_n^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1/2}|| & = \Big[\mathbf{b}_n^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\mathbf{X}_i\mathbf{b}_n\Big]^{1/2} \\
& \leq \lambda_{\max}^{1/2}(\overline{\mathbf{R}}^{-1})\lambda_{\max}^{1/2}(\mathbf{A}_i(\boldsymbol{\beta}_n))||\mathbf{X}_i\mathbf{b}_n||,
\end{aligned}
$$

and

$$
\begin{aligned}
& ||\overline{\mathbf{R}}^{-1/2}[\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0}) - \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})]\mathbf{X}_i\mathbf{b}_n|| \\
\leq & \lambda_{\max}^{1/2}(\overline{\mathbf{R}}^{-1})\lambda_{\max}^{1/2}([\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0}) - \mathbf{A}_i(\boldsymbol{\beta}_n)]^2)||\mathbf{X}_i\mathbf{b}_n||.
\end{aligned}
$$

Thus

$$
\begin{aligned}
I_{n1} & \leq C\max_{i,j}\Big|\mathbf{A}_{ij}^{1/2}(\boldsymbol{\beta}_{n0}) - \mathbf{A}_{ij}^{1/2}(\boldsymbol{\beta}_n)\Big|\sum_{i=1}^n ||\mathbf{X}_i\mathbf{b}_n||^2 \\
& \leq C\max_{i,j}||\mathbf{X}_{ij}|| \cdot ||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}|| \cdot \lambda_{\max}(\sum_{i=1}^n \mathbf{X}_i^T\mathbf{X}_i)||\mathbf{b}_n||^2.
\end{aligned}
$$

Therefore,

$$
\sup_{||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}|| \leq \Delta\sqrt{\frac{p_n}{n}}} \sup_{||\mathbf{b}_n||=1} I_{n1} = O_p(\sqrt{n}p_n).
$$

Similarly, $\sup_{||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}|| \le \Delta\sqrt{\frac{p_n}{n}}} \sup_{||\mathbf{b}_n||=1} I_{n2} = O_p(\sqrt{n}p_n)$ and this proves the lemma. □

APPENDIX B

Sketch of proof of (3.11) in Remark 6. A nice and clean proof can be given for the random design case and relies on the following high-dimensional CLT result of Portnoy (1988, Section 4):

Theorem B.1. *Let $\mathbf{W}_1, \ldots, \mathbf{W}_n$ be p_n-dimensional independent and identically distributed random vectors with $E(\mathbf{W}_1) = \mathbf{0}$, $Cov(\mathbf{W}_1) = \mathbf{I}$ and $E(W_{1j}^6) \le B \le \infty$, for $j = 1, \ldots, p_n$, where W_{1j} denotes the jth element of $\mathbf{W}_1$. Then, if $p_n/n \to 0$,*

$$\frac{n^{-1}||\sum_{i=1}^n \mathbf{W}_i||^2 - p_n}{\sqrt{2p_n}} \to N(0,1)$$

in distribution.

The above CLT was proved by Portnoy using a martingale technique, which is of considerable interest on its own. Now we consider the random design setting where $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im_i})^T$ is a random matrix. Assuming that $E(\mathbf{X}_i^T\mathbf{X}_i)$ has bounded eigenvalues, it can be shown that the results we have in Theorem 3.6 and Theorem 3.8 and the associated lemmas still hold with $\overline{\mathbf{M}}_n(\boldsymbol{\beta}_n)$ now denoting $n\mathrm{E}(\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{R}_0\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\mathbf{X}_i)$. From the proof of Theorem 3.8, we have the following expression for $\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*$:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^* &= \overline{\mathbf{H}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*) - \overline{\mathbf{H}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)[\mathbf{D}_n(\widetilde{\boldsymbol{\beta}}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_{n0}^*)](\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*)\\ &\quad -\overline{\mathbf{H}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)[\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*) - \mathbf{S}_n(\boldsymbol{\beta}_{n0}^*)],\end{aligned}$$

where $\widetilde{\boldsymbol{\beta}}_n$ is between $\boldsymbol{\beta}_{n0}^*$ and $\widehat{\boldsymbol{\beta}}_n$. Using the above expression, we have

$$\begin{aligned}&\frac{(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*)^T\boldsymbol{\Sigma}_n^{-1}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*) - p_n}{\sqrt{2p_n}}\\ &= \frac{\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*)^T\overline{\mathbf{H}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)\boldsymbol{\Sigma}_n^{-1}\overline{\mathbf{H}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*) - p_n}{\sqrt{2p_n}}\\ &\quad + \text{eight other terms}.\end{aligned} \tag{B.1}$$

Note that

$$\begin{aligned}\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*)^T\overline{\mathbf{H}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)\boldsymbol{\Sigma}_n^{-1}\overline{\mathbf{H}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*) &= \overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*)^T\overline{\mathbf{M}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*)\\ &= ||\overline{\mathbf{M}}_n^{-1/2}(\boldsymbol{\beta}_{n0}^*)\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*)||^2.\end{aligned}$$

We can write $\overline{\mathbf{M}}_n^{-1/2}(\boldsymbol{\beta}_{n0}^*)\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*) = n^{-1/2}\sum_{i=1}^n \mathbf{W}_i$, where

$$\mathbf{W}_i = [n^{-1}\overline{\mathbf{M}}_n(\boldsymbol{\beta}_{n0}^*)]^{-1/2}\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0}^*)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-1/2}(\boldsymbol{\beta}_{n0}^*)(\mathbf{Y}_i - \boldsymbol{\pi}_i(\boldsymbol{\beta}_{n0}^*)).$$

Note that $\mathrm{E}(\mathbf{W}_i) = \mathbf{0}$ and $\mathrm{Var}(\mathbf{W}_i) = \mathbf{I}$, and if we assume $E((\mathbf{X}_i^T\mathbf{X}_i)^3)$ has the largest eigenvalue bounded, then the $E(W_{1j}^6) \leq B$ for some finite positive constant B. Thus applying Portnoy's CLT, the first term on the right side of (B.1) converges to $N(0,1)$ in distribution. All the other eight terms on the right side of (B.1) are asymptotically negligible. As an example, we consider the following typical term:

$$\begin{aligned}
& I_n \\
= \; & (2p_n)^{-1}\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*)^T\overline{\mathbf{H}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)\boldsymbol{\Sigma}_n^{-1}\overline{\mathbf{H}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)[\mathbf{D}_n(\widetilde{\boldsymbol{\beta}}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_{n0}^*)](\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*) \\
= \; & (2p_n)^{-1}\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*)^T\overline{\mathbf{M}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)[\mathbf{D}_n(\widetilde{\boldsymbol{\beta}}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_{n0}^*)](\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*).
\end{aligned}$$

By Cauchy-Schwarz inequality,

$$\begin{aligned}
I_n \;\leq\; & (2p_n)^{-1}||\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*)^T\overline{\mathbf{M}}_n^{-1}(\boldsymbol{\beta}_{n0}^*)[\mathbf{D}_n(\widetilde{\boldsymbol{\beta}}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_{n0}^*)]|| \cdot ||\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*|| \\
\leq\; & (2p_n)^{-1}||\overline{\mathbf{M}}_n^{-1/2}(\boldsymbol{\beta}_{n0}^*)\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}^*)|| \cdot ||\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*|| \cdot \lambda_{min}^{-1/2}(\overline{\mathbf{M}}_n(\boldsymbol{\beta}_{n0}^*)) \\
& \max(|\lambda_{min}(\mathbf{D}_n(\widetilde{\boldsymbol{\beta}}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_{n0}^*))|, |\lambda_{max}(\mathbf{D}_n(\widetilde{\boldsymbol{\beta}}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_{n0}^*))|) \\
=\; & (2p_n)^{-1}O_p(\sqrt{p_n})O_p(\sqrt{p_n/n})O_p(n^{-1/2})O_p(\sqrt{n}p_n) \\
=\; & O_p(p_n^{3/2}n^{-1/2}) = o_p(1),
\end{aligned}$$

where it follows from Lemmas 3.3-3.5 that

$$\max(|\lambda_{min}(\mathbf{D}_n(\widetilde{\boldsymbol{\beta}}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_{n0}^*))|, |\lambda_{max}(\mathbf{D}_n(\widetilde{\boldsymbol{\beta}}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_{n0}^*))|) = O_p(\sqrt{n}p_n).$$

Finally, the proof will be completed by showing

$$(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*)^T(\widehat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1})(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*)/\sqrt{2p_n} = o_p(1). \tag{B.2}$$

Note that

$$\begin{aligned}
& |p_n^{-1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*)^T(\widehat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1})(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*)| \\
= \; & |p_n^{-1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*)^T\widehat{\boldsymbol{\Sigma}}_n^{-1}(\boldsymbol{\Sigma}_n - \widehat{\boldsymbol{\Sigma}}_n)\boldsymbol{\Sigma}_n^{-1}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*)| \\
\leq \; & p_n^{-1/2}||\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}^*||^2\lambda_{min}^{-1}(\widehat{\boldsymbol{\Sigma}}_n)\lambda_{min}^{-1}(\boldsymbol{\Sigma}_n) \\
& \cdot \max(|\lambda_{min}(\boldsymbol{\Sigma}_n - \widehat{\boldsymbol{\Sigma}}_n)|, |\lambda_{max}(\boldsymbol{\Sigma}_n - \widehat{\boldsymbol{\Sigma}}_n)|).
\end{aligned}$$

In the proof of Corollary 3.11, we have shown that $\lambda_{min}(\boldsymbol{\Sigma}_n) \geq O_p(n^{-1})$. Similarly, we can show $\lambda_{min}(\widehat{\boldsymbol{\Sigma}}_n) \geq O_p(n^{-1})$. And contained in the proof is the result that $\max(|\lambda_{min}(\boldsymbol{\Sigma}_n - \widehat{\boldsymbol{\Sigma}}_n)|, |\lambda_{max}(\boldsymbol{\Sigma}_n - \widehat{\boldsymbol{\Sigma}}_n)|) = o_p(n^{-1})$. Putting these together verifies (B.2). $\square$

APPENDIX C

Sketch of proof of Theorem 5.1. The proof will highlight the places where the proof differs from the binary case. We denote the generalized estimating equation by

$$\mathbf{S}_n = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\widehat{\mathbf{R}}^{-1}\mathbf{A}_i^{-1/2}(\boldsymbol{\beta}_n)(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_n)).$$

Define $\overline{\mathbf{S}}_n(\boldsymbol{\beta}_n)$, $\overline{\mathbf{M}}_n(\boldsymbol{\beta}_n)$, $\mathbf{D}_n(\boldsymbol{\beta}_n)$ and $\overline{\mathbf{D}}_n(\boldsymbol{\beta}_n)$ similarly as in the clustered binary data case.

Lemma C.1.

$$\overline{\mathbf{D}}_n(\boldsymbol{\beta}_n) = \overline{\mathbf{H}}_n(\boldsymbol{\beta}_n) + \overline{\mathbf{E}}_n(\boldsymbol{\beta}_n) + \overline{\mathbf{G}}_n(\boldsymbol{\beta}_n), \tag{C.1}$$

where

$$\begin{aligned}
\overline{\mathbf{H}}_n(\boldsymbol{\beta}_n) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\mathbf{X}_i, \\
\overline{\mathbf{E}}_n(\boldsymbol{\beta}_n) &= \frac{1}{2}\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_n)\mathbf{C}_i(\boldsymbol{\beta}_n)\mathbf{F}_i(\boldsymbol{\beta}_n)\mathbf{X}_i, \\
\overline{\mathbf{G}}_n(\boldsymbol{\beta}_n) &= -\frac{1}{2}\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\mathbf{F}_i(\boldsymbol{\beta}_n)\mathbf{J}_i(\boldsymbol{\beta}_n)\mathbf{X}_i,
\end{aligned}$$

with

$$\begin{aligned}
\mathbf{C}_i(\boldsymbol{\beta}_n) &= diag(Y_{i1} - \mu_{i1}(\boldsymbol{\beta}_n), \ldots, Y_{im} - \mu_{im}(\boldsymbol{\beta}_n)), \\
\mathbf{F}_i(\boldsymbol{\beta}_n) &= diag(\ddot{\mu}(\mathbf{X}_{i1}^T\boldsymbol{\beta}_n), \ldots, \ddot{\mu}(\mathbf{X}_{im}^T\boldsymbol{\beta}_n)), \\
\mathbf{J}_i(\boldsymbol{\beta}_n) &= diag(\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-1/2}(\boldsymbol{\beta}_n)(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_n))).
\end{aligned}$$

In the above, for $\mathbf{a} = (a_1, \ldots, a_m)^T$, diag$(a_1, \ldots, a_m)$ and diag$(\mathbf{a})$ both denote an $m \times m$ diagonal matrix with diagonal entries $a_1, \ldots, a_m$.

Remark. The above decomposition has a slightly different form but is equivalent to that given in Xie and Yang (2003), which can be derived by straightforward but tedious calculation of matrix differentiation.

Lemma C.2. *Assume conditions (A1)–(A6), if* $n^{-1}p_n^2 = o(1)$, *then*

$$||\mathbf{S}_n(\boldsymbol{\beta}_{n0}) - \overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0})|| = O_p(p_n). \tag{C.2}$$

Furthermore, for $\mathbf{b}_n \in R^{p_n}$*, we have*

$$\text{(C.3)} \quad \sup_{||\boldsymbol{\beta}_n-\boldsymbol{\beta}_{n0}||\le\Delta\sqrt{\frac{p_n}{n}}} \sup_{||\mathbf{b}_n||=1} \left|\mathbf{b}_n^T[\mathbf{D}_n(\boldsymbol{\beta}_n) - \overline{\mathbf{D}}_n(\boldsymbol{\beta}_n)]\mathbf{b}_n\right| = O_p(\sqrt{np_n}).$$

$$\text{(C.4)} \quad \sup_{||\boldsymbol{\beta}_n-\boldsymbol{\beta}_{n0}||\le\Delta\sqrt{\frac{p_n}{n}}} \sup_{||\mathbf{b}_n||=1} \left|\mathbf{b}_n^T[\overline{\mathbf{D}}_n(\boldsymbol{\beta}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_n)]\mathbf{b}_n\right| = O_p(\sqrt{n}p_n).$$

$$\text{(C.5)} \quad \sup_{||\boldsymbol{\beta}_n-\boldsymbol{\beta}_{n0}||\le\Delta\sqrt{\frac{p_n}{n}}} \sup_{||\mathbf{b}_n||=1} \left|\mathbf{b}_n^T[\overline{\mathbf{H}}_n(\boldsymbol{\beta}_n) - \overline{\mathbf{H}}_n(\boldsymbol{\beta}_{n0})]\mathbf{b}_n\right| = O_p(\sqrt{n}p_n).$$

PROOF. We can prove (C.2) and (C.5) similarly as in the binary case. To prove (C.3), since $\mathbf{H}_n(\boldsymbol{\beta}_n)$ and $\overline{\mathbf{H}}_n(\boldsymbol{\beta}_n)$ have the same expressions as in the binary case, it suffices to show that

$$\text{(C.6)} \quad \sup_{||\boldsymbol{\beta}_n-\boldsymbol{\beta}_{n0}||\le\Delta\sqrt{\frac{p_n}{n}}} \sup_{||\mathbf{b}_n||=1} \left|\mathbf{b}_n^T[\mathbf{E}_n(\boldsymbol{\beta}_n) - \overline{\mathbf{E}}_n(\boldsymbol{\beta}_n)]\mathbf{b}_n\right| = O_p(\sqrt{np_n}).$$

$$\text{(C.7)} \quad \sup_{||\boldsymbol{\beta}_n-\boldsymbol{\beta}_{n0}||\le\Delta\sqrt{\frac{p_n}{n}}} \sup_{||\mathbf{b}_n||=1} \left|\mathbf{b}_n^T[\mathbf{G}_n(\boldsymbol{\beta}_n) - \overline{\mathbf{G}}_n(\boldsymbol{\beta}_n)]\mathbf{b}_n\right| = O_p(\sqrt{np_n}).$$

Let $\mathbf{C}_{i1}(\boldsymbol{\beta}_n) = \mathrm{diag}(\mu_{i1}(\boldsymbol{\beta}_{n0})-\mu_{i1}(\boldsymbol{\beta}_n), \ldots, \mu_{im}(\boldsymbol{\beta}_{n0})-\mu_{im}(\boldsymbol{\beta}_n))$ and $\mathbf{C}_{i2}(\boldsymbol{\beta}_{n0}) = \mathrm{diag}(Y_{i1} - \mu_{i1}(\boldsymbol{\beta}_{n0}), \ldots, Y_{im} - \mu_{im}(\boldsymbol{\beta}_{n0}))$. Then

$$\begin{aligned}
&\left|\mathbf{b}_n^T[\mathbf{E}_n(\boldsymbol{\beta}_n) - \overline{\mathbf{E}}_n(\boldsymbol{\beta}_n)]\mathbf{b}_n\right| \\
\le\ & \frac{1}{2}\left|\sum_{i=1}^n b_n^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_n)\mathbf{C}_{i1}(\boldsymbol{\beta}_n)\mathbf{F}_i(\boldsymbol{\beta}_n)\mathbf{X}_i b_n\right| \\
&+\frac{1}{2}\left|\sum_{i=1}^n b_n^T\mathbf{X}_i^T[\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n) - \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})](\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_n)\mathbf{C}_{i2}(\boldsymbol{\beta}_{n0})\mathbf{F}_i(\boldsymbol{\beta}_n)\mathbf{X}_i b_n\right| \\
&+\frac{1}{2}\left|\sum_{i=1}^n b_n^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})[\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_n) - \mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})]\mathbf{C}_{i2}(\boldsymbol{\beta}_{n0})\mathbf{F}_i(\boldsymbol{\beta}_n)\mathbf{X}_i b_n\right| \\
&+\left|\frac{1}{2}\sum_{i=1}^n b_n^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})\mathbf{C}_{i2}(\boldsymbol{\beta}_{n0})[\mathbf{F}_i(\boldsymbol{\beta}_n) - \mathbf{F}_i(\boldsymbol{\beta}_{n0})]\mathbf{X}_i b_n\right| \\
&+\left|\frac{1}{2}\sum_{i=1}^n b_n^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})\mathbf{C}_{i2}(\boldsymbol{\beta}_{n0})\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{X}_i b_n\right| \\
\triangleq\ & \sum_{k=1}^4 J_{nk}(\boldsymbol{\beta}_n) + J_{n5}(\boldsymbol{\beta}_{n0}).
\end{aligned}$$

By Cauchy-Schwarz inequality for matrices with Frobenius norm,

$$\begin{aligned} J_{n1} &\leq \frac{1}{2}\sum_{i=1}^{n}||\mathbf{X}_i b_n||^2 \cdot ||\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1}|| \cdot \frac{\max_{1\leq j\leq m}\sigma_{ij}(\boldsymbol{\beta}_n)}{\min_{1\leq j\leq m}\sigma_{ij}^3(\boldsymbol{\beta}_n)} \max_{1\leq j\leq m}|\dot{\mu}(\mathbf{X}_{i1}^T\tilde{\boldsymbol{\beta}}_n)| \\ &\quad \max_{1\leq j\leq m}|\ddot{\mu}(\mathbf{X}_{i1}^T\tilde{\boldsymbol{\beta}}_n)| \cdot ||\boldsymbol{\beta}_{n0} - \boldsymbol{\beta}_n||, \end{aligned}$$

where $\tilde{\boldsymbol{\beta}}_n$ is between $\boldsymbol{\beta}_n$ and $\boldsymbol{\beta}_{n0}$. Thus

$$\begin{aligned} &\sup_{||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}||\leq\Delta\sqrt{\frac{p_n}{n}}} \sup_{||\mathbf{b}_n||=1} J_{n1}(\boldsymbol{\beta}_n) \\ \leq\ & CO_p(\sqrt{p_n/n})O_p(\sqrt{p_n/n})\lambda_{max}(\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{X}_i) = O_p(\sqrt{np_n}). \end{aligned}$$

Similarly, we can show that $\sup_{||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}||\leq\Delta\sqrt{\frac{p_n}{n}}} \sup_{||\mathbf{b}_n||=1} J_{nk}(\boldsymbol{\beta}_n) = O_p(\sqrt{np_n})$, for $k = 2, 3, 4$. To evaluate $J_{n5}(\boldsymbol{\beta}_{n0})$, note that $\sup_{||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}||\leq\Delta\sqrt{\frac{p_n}{n}}} \sup_{||\mathbf{b}_n||=1} J_{n5}(\boldsymbol{\beta}_{n0}) = \sup_{||\mathbf{b}_n||=1} J_{n5}(\boldsymbol{\beta}_{n0}) \leq ||J_{n5}(\boldsymbol{\beta}_{n0})||)$. Similarly as in the proof of Lemma 3.4, its sufficient to show that $E(||J_{n5}(\boldsymbol{\beta}_{n0})||^2) = O(np_n)$. We have

$$\begin{aligned} & E[||J_{n5}(\boldsymbol{\beta}_{n0})||^2] \\ =\ & \mathrm{trace}\left[E\left(J_{n5}(\boldsymbol{\beta}_{n0})J_{n5}^T(\boldsymbol{\beta}_{n0})\right)\right] \\ =\ & \frac{1}{4}\sum_{i=1}^{n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m} E[\epsilon_{ij_1}(\boldsymbol{\beta}_{n0})\epsilon_{ij_2}(\boldsymbol{\beta}_{n0})]\mathrm{trace}\Big[\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0}) \\ & \mathbf{e}_{j_1}\mathbf{e}_{j_1}^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{X}_i\mathbf{X}_i^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{e}_{j_2}\mathbf{e}_{j_2}^T\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})\mathbf{X}_i\Big] \\ \leq\ & C\sum_{i=1}^{n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\mathbf{e}_{j_1}^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{X}_i\mathbf{X}_i^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{e}_{j_2}\mathbf{e}_{j_2}^T\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1}) \\ & \cdot vA_i^{1/2}(\boldsymbol{\beta}_{n0})\mathbf{X}_i\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})\mathbf{e}_{j_1} \\ \leq\ & C\sum_{i=1}^{n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}||\mathbf{e}_{j_1}^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{X}_i|| \cdot ||\mathbf{X}_i^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{e}_{j_2}|| \\ & \cdot||\mathbf{e}_{j_2}^T\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})\mathbf{X}_i|| \\ & \cdot||\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})\mathbf{e}_{j_1}||. \end{aligned}$$

Note that $||\mathbf{X}_i^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{e}_{j_2}|| \leq C||\mathbf{X}_{ij_1}||$ and $||\mathbf{X}_i^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{e}_{j_2}|| \leq C||\mathbf{X}_{ij_2}||$. Further, $||\mathbf{e}_{j_2}^T\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})\mathbf{X}_i|| \leq C||\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1}||(\mathrm{trace}(\mathbf{X}_i\mathbf{X}_i^T))^{1/2}$ and $||\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})(\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1})\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})\mathbf{e}_{j_1}|| \leq C||\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1}||(\mathrm{trace}(\mathbf{X}_i\mathbf{X}_i^T))^{1/2}$

Thus

$$
\begin{aligned}
& E[||J_{n5}(\boldsymbol{\beta}_{n0})||^2] \\
\leq\ & C\sum_{i=1}^{n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m} ||\mathbf{X}_{ij_1}|| \cdot ||\mathbf{X}_{ij_2}|| \cdot ||\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1}||^2 \cdot \mathrm{trace}(\mathbf{X}_i\mathbf{X}_i^T) \\
\leq\ & C \cdot \max_{i,j} ||\mathbf{X}_{ij}||^2 ||\widehat{\mathbf{R}}^{-1} - \overline{\mathbf{R}}^{-1}||^2 \mathrm{trace}(\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T) = O(p_n^3) = o(np_n)
\end{aligned}
$$

by assumption A1. This implies that $\sup_{||\mathbf{b}_n||=1} |\mathbf{b}_n^T J_{n5}(\boldsymbol{\beta}_{n0})\mathbf{b}_n| = O_p(\sqrt{np_n})$. Therefore, (C.6) is established. Similarly, we can prove (C.7) and hence verify (C.3). To prove (C.4), it is sufficient to show that

$$
\text{(C.8)} \qquad \sup_{||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}|| \leq \Delta\sqrt{\frac{p_n}{n}}} \ \sup_{||\mathbf{b}_n||=1} \left|\mathbf{b}_n^T \overline{\mathbf{E}}_n(\boldsymbol{\beta}_n)\mathbf{b}_n\right| = O_p(\sqrt{n}p_n),
$$

$$
\text{(C.9)} \qquad \sup_{||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}|| \leq \Delta\sqrt{\frac{p_n}{n}}} \ \sup_{||\mathbf{b}_n||=1} \left|\mathbf{b}_n^T \overline{\mathbf{G}}_n(\boldsymbol{\beta}_n)\mathbf{b}_n\right| = O_p(\sqrt{n}p_n).
$$

To prove (C.8), note that we have the following decomposition of $\overline{\mathbf{E}}_n(\boldsymbol{\beta}_n)$:

$$
\begin{aligned}
& \overline{\mathbf{E}}_n(\boldsymbol{\beta}_n) \\
=\ & \frac{1}{2}\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})\mathrm{diag}(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_{n0}))\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{X}_i \\
& + \frac{1}{2}\sum_{i=1}^{n} \mathbf{X}_i^T [\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n) - \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})]\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})\mathrm{diag}(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_{n0}))\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{X}_i \\
& + \frac{1}{2}\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}[\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_n) - \mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})]\mathrm{diag}(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_{n0}))\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{X}_i \\
& + \frac{1}{2}\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_n)\mathrm{diag}(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_{n0}))[\mathbf{F}_i(\boldsymbol{\beta}_n) - \mathbf{F}_i(\boldsymbol{\beta}_{n0})]\mathbf{X}_i \\
& + \frac{1}{2}\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_n)\mathrm{diag}(\boldsymbol{\mu}_i(\boldsymbol{\beta}_{n0}) - \boldsymbol{\mu}_i(\boldsymbol{\beta}_n))\mathbf{F}_i(\boldsymbol{\beta}_n)\mathbf{X}_i \\
\triangleq\ & \overline{\mathbf{E}}_{1n}(\boldsymbol{\beta}_{n0}) + \sum_{k=2}^{5}\overline{\mathbf{E}}_{kn}(\boldsymbol{\beta}_n).
\end{aligned}
$$

Using the same techniques as above, we can prove that $||\overline{\mathbf{E}}_{1n}(\boldsymbol{\beta}_{n0})|| =$

$O_p(\sqrt{n}p_n)$, thus $\sup_{||\mathbf{b}_n||=1}|\mathbf{b}_n^T\overline{\mathbf{E}}_{1n}(\boldsymbol{\beta}_{n0}\mathbf{b}_n| = O_p(\sqrt{n}p_n)$. Next,

$$
\begin{aligned}
&\left|\mathbf{b}_n^T\overline{\mathbf{E}}_{2n}(\boldsymbol{\beta}_{n0})\mathbf{b}_n\right| \\
= \; & \frac{1}{2}\Big|\sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij}-\mu_{ij}(\boldsymbol{\beta}_{n0}))\mathbf{b}_n^T\mathbf{X}_i^T[\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)-\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})]\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0}) \\
& \mathbf{e}_j\mathbf{e}_j^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{X}_i\mathbf{b}_n\Big| \\
\le \; & \frac{1}{2}\Big[\sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij}-\mu_{ij}(\boldsymbol{\beta}_{n0}))^2\Big]^{1/2}\Big[\sum_{i=1}^{n}\sum_{j=1}^{m}(\mathbf{b}_n^T\mathbf{X}_i^T[\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)-\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})] \\
& \overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})\mathbf{e}_j\mathbf{e}_j^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{X}_i\mathbf{b}_n)^2\Big]^{1/2} \\
= \; & O_p(\sqrt{n})\Big[\sum_{i=1}^{n}\sum_{j=1}^{m}||\mathbf{b}_n^T\mathbf{X}_i^T[\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)-\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})]\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_{n0})\mathbf{e}_j||^2 \\
& \cdot||\mathbf{e}_j^T\mathbf{F}_i(\boldsymbol{\beta}_{n0})\mathbf{X}_i\mathbf{b}_n||^2\Big]^{1/2} \\
\le \; & O_p(\sqrt{n})\Big[\sum_{i=1}^{n}\sum_{j=1}^{m}||\mathbf{X}_i\mathbf{b}_n||^2\max_{1\le j\le m}(\sigma_j^2(\boldsymbol{\beta}_n)-\sigma_j^2(\boldsymbol{\beta}_{n0}))^2\lambda_{\max}(\overline{\mathbf{R}}^{-1}) \\
& (\min_{1\le j}\sigma(\boldsymbol{\beta}_{n0}))^{-3}\max_{1\le j\le m}(\ddot{\mu}^2(\mathbf{X}_{ij}^T\boldsymbol{\beta}_{n0}))||\mathbf{X}_{ij}||^2\Big]^{1/2}
\end{aligned}
$$

Thus

$$
\begin{aligned}
& \sup_{||\boldsymbol{\beta}_n-\boldsymbol{\beta}_{n0}||\le\Delta\sqrt{\frac{p_n}{n}}}\;\sup_{||\mathbf{b}_n||=1}\left|\mathbf{b}_n^T\overline{\mathbf{E}}_{2n}(\boldsymbol{\beta}_{n0})\mathbf{b}_n\right| \\
\le \; & O_p(\sqrt{n})\Big[O_p(p_n/n)O(n)O(P_n)\Big]^{1/2} = O_p(\sqrt{n}p_n).
\end{aligned}
$$

Similarly, we can show that $\sup_{||\boldsymbol{\beta}_n-\boldsymbol{\beta}_{n0}||\le\Delta\sqrt{\frac{p_n}{n}}}\sup_{||\mathbf{b}_n||=1}\left|\mathbf{b}_n^T\overline{\mathbf{E}}_{in}(\boldsymbol{\beta}_{n0})\mathbf{b}_n\right| = O_p(\sqrt{n}p_n)$, $i=3,4,5$. This proves (C.8), and (C.9) can be proved the same way. Together this verifies (C.4). □

Proof of Theorem 5.1. The existence and consistency of $\widehat{\boldsymbol{\beta}}_n$ can be established similarly as in the binary case by applying Lemma C.1 and Lemma C.2. To prove asymptotic normality, we first show that $\forall\ \boldsymbol{\alpha}_n \in R^{p_n}$ such that $||\boldsymbol{\alpha}_n||=1$, we have

$$
\boldsymbol{\alpha}_n^T\overline{\mathbf{M}}_n^{-1/2}(\boldsymbol{\beta}_{n0})\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}) \to N(0,1) \quad \text{in distribution.} \tag{C.10}
$$

We write $\boldsymbol{\alpha}_n^T\overline{\mathbf{M}}_n^{-1/2}(\boldsymbol{\beta}_{n0})\overline{\mathbf{S}}_n(\boldsymbol{\beta}_{n0}) = \sum_{i=1}^n Z_{ni}$, where

$$
Z_{ni} = \boldsymbol{\alpha}_n^T\overline{\mathbf{M}}_n^{-1/2}(\boldsymbol{\beta}_{n0})\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})\overline{\mathbf{R}}^{-1}\boldsymbol{\epsilon}_i(\boldsymbol{\beta}_{n0}),
$$

with $\boldsymbol{\epsilon}_i(\boldsymbol{\beta}_{n0}) = \mathbf{A}_i^{-1/2}(\boldsymbol{\beta}_{n0})(\mathbf{Y}_i - \boldsymbol{\pi}(\boldsymbol{\beta}_{n0}))$. Then $\mathrm{E}(Z_{ni}) = 0$, $\mathrm{Var}(\sum_{i=1}^n Z_{ni}) = 1$. As in the proof of Lemma 3.7, by Cauchy-Schwarz inequality, we have $Z_{ni}^2 \leq C\gamma_{ni}||\boldsymbol{\epsilon}_i(\boldsymbol{\beta}_{n0})||^2$, where $\gamma_{ni} = \boldsymbol{\alpha}_n^T \overline{\mathbf{M}}_n^{-1/2}(\boldsymbol{\beta}_{n0})\mathbf{X}_i^T\mathbf{X}_i\overline{\mathbf{M}}_n^{-1/2}(\boldsymbol{\beta}_{n0})\boldsymbol{\alpha}_n$ and $\max_{1\leq i\leq n}\gamma_{ni} \leq O(n^{-1}p_n) = o(1)$. To prove (C.10), we check Lyapunov condition. For arbitrary $\epsilon > 0$, we want to show

$$\sum_{i=1}^n E\left(|Z_{ni}|^{2+\delta}\right) \to 0, \tag{C.11}$$

as $n \to \infty$. Note that

$$\begin{aligned}
\sum_{i=1}^n E\left(|Z_{ni}|^{2+\delta}\right) &\leq \sum_{i=1}^n E(C^{1+\delta/2}\gamma_{ni}^{1+\delta/2}||\boldsymbol{\epsilon}_i(\boldsymbol{\beta}_{n0})||^{2+\delta}) \\
&\leq C(\max_{1\leq i\leq n}\gamma_{ni})^{\delta/2}\sum_{i=1}^n \gamma_{ni} \\
&\leq C(\max_{1\leq i\leq n}\gamma_{ni})^{\delta/2}\sum_{i=1}^n \boldsymbol{\alpha}_n^T \overline{\mathbf{M}}_n^{-1/2}(\boldsymbol{\beta}_{n0})\mathbf{X}_i^T\mathbf{X}_i\overline{\mathbf{M}}_n^{-1/2}(\boldsymbol{\beta}_{n0})\boldsymbol{\alpha}_n \\
&\leq C(\max_{1\leq i\leq n}\gamma_{ni})^{\delta/2}\lambda_{\max}(\sum_{i=1}^n \mathbf{X}_i^T\mathbf{X}_i)\lambda_{\min}^{-1}(\overline{\mathbf{M}}_n(\boldsymbol{\beta}_{n0})) \\
&= o(1)O(n)O(n^{-1}) = o(1).
\end{aligned}$$

Thus (C.10) is proved. The asymptotic normality can be proved similarly as for Theorem 3.8. □

School of Statistics
University of Minnesota,
224 Church Street, SE
Minneapolis, MN 55455
E-mail: lan@stat.umn.edu